IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	Weissman	Art Unit:	2166
Serial No.:	10/748,505	Examiner:	Sangwoo Ahn
Filed:	December 30, 2003	Conf. No.:	7264
Title:	METHODS AND SYSTEMS FOR COMPRESSING INDICES		

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REPLY BRIEF

Pursuant to 37 C.F.R. § 41.41, Applicant responds to the Examiner's Answer mailed January 9, 2008 as follows:

At page 11, line 8–13, the Examiner's Answer contends that "recording associated articles in two cells is not distinguishable over recording them in a single cell, since it is evident that both articles are associated with a single item." This contention allegedly supports the Examiner's "disagreement" with Applicant's contention that each entry in Chaudhuri's FIG. 4 includes a single column ID value.

Applicant respectfully disagrees with the contention that recording associated articles in two cells "is not distinguishable over" recording them in a single cell. Indeed, this contention is particularly disturbing in the context of the compression of entries in an index and does not set forth a reasonable basis for disagreement with Applicant's contention that each entry in Chaudhuri's FIG. 4 includes a single column ID value.

The present claims all deal with the compression of item entries. For example, claim 1 recites that a second item entry selected from an inverted index is compressed into a first item entry selected from the same inverted index. The compression of item entries inherently requires that the recording of information in multiple cells and entries be distinguished from the recording of information in a single cell or a single entry. The Examiner's bald contention that an uncompressed state (with data recorded in first and second item entries) "is not distinguishable over" a compressed state (with data recorded in a single item entry) is thus uniquely pernicious in the present context. Indeed, it amounts to a contention that any such compression of cells and entries is somehow excluded from patentability since the initial state and the final state are allegedly not "distinguishable."

This is manifestly improper. For example, both Applicant's specification and the cited Chaudhuri reference describe that recording associated articles in two cells or entries is "distinguishable over" recording them in a single cells or entries. For example, as discussed in Applicant's specification, inverted indices that characterize the World Wide Web and enterprise network systems are becoming extremely large. *See, e.g., specification*, para. [0005]. The large size of these inverted indices occupies a great deal of memory space. *Id.* By compressing entries, benefits such as memory savings and processing time savings can be achieved. *See, e.g., specification*, para. [0031]. Chaudhuri describes that the size of his hash tables can be reduced using a loss-less compression technique and that such reductions in size are beneficial. *See, e.g., Chaudhuri*, para. [0035]. Accordingly, both Applicant's specification and the cited Chaudhuri reference consider the recording of information in multiple cells and entries to be "distinguishable over" recording information in a single cell and entry.

Indeed, it appears that the Examiner has declared that compressing cells and entries is somehow excluded from patentability based on the initial state and the final state allegedly being indistinguishable. However, Applicant is not aware of any authority which excludes the compression of cells and entries from being patentable. If the Examiner is aware of any such authority, Applicant respectfully requests that it be provided.

Also, even if recording associated articles in two cells or item entries were "not distinguishable over" recording them in a single cell or item entry (a contention with which Applicant does not agree), this does not disparage Applicant's contention that each entry in Chaudhuri's FIG. 4 includes a single column ID. Applicant's characterization of Chaudhuri's FIG. 4 is entirely correct. The Examiner's position does not change the content of Chaudhuri.

At page 11, line 13–15, the Examiner's Answer contends that "the term 'entry' has a broad range of interpretation, thus its interpretation is not only restricted to a single cell/row/column in a table." **At page 12, row 1 and row 3**, the Examiner's Answer contends that multiple rows in Chaudhuri's FIG. 4 constitute a (single) entry.

Applicant disagrees with the contention that multiple rows in a table are reasonably considered to be a single entry and submits that there is no reasonable basis for this contention. To begin with, applicant would like to point out that the claims all relate to entries from an inverted index. Any "broad range of interpretation" based on usage of the term "entry" outside of this context (such as, e.g., "an entry in the Kentucky Derby" or "an entry into the castle") would appear to be irrelevant.

Further, the Examiner's Answer has not developed how the allegedly "broad range of interpretation" supports the contention that multiple rows in a table are reasonably considered to be a single entry. Indeed, neither the Examiner's Answer nor the previous correspondence in this application has established any basis for believing that the term "entry," in the context of an inverted index, can encompass multiple rows in Chaudhuri's table.

Instead, the references of record in this application are understood to use the term "entry" in a manner inconsistent with the contention that multiple rows in Chaudhuri's table constitute a single "entry." For example, U.S. Patent No. 5,745,898 to Burrows (hereinafter "Burrows I") and U.S. Patent No. 5,765,168 to Burrows (hereinafter "Burrows II") also use the term "entry" in the context on an inverted index without the "broad range of interpretation" allegedly discerned by the Examiner. *See, e.g., Burrows II*, FIGS. 6-8 and the written description thereof. There is simply no reason to belief that those of ordinary skill, when armed with the specification and an understanding of the context of the claims, would not be able to recognize that multiple rows in the table of Chaudhuri's FIG. 4 are not a single entry.

Accordingly, Applicant submits that the interpretation of the term "entry" relied upon by the Examiner's Answer to encompass multiple rows in a single table is incorrect and unreasonable. An anticipation rejection cannot be supported on this basis. If the Examiner persists in the contention that the "broad range of interpretation" of the term "entry," in the context of an inverted index, encompasses multiple rows in a single table, applicant respectfully requests that some iota of evidentiary support for this contention be provided.

At page 11, line 15 – page 12, line 2, the Examiner's Answer contends that:

> "[t]he first item entry and the second item are in fact different in Chaudhuri, since each of the hash values V2 and V3 represents a disparate keyword. While items V2 and V3 are different hash values, their contents (in other words, associated articles) are the same, thus resulting in compression of those two contents by having a single value or data which represents the articles in common. This step could be interpreted as equivalent to the present claim's 'compressing the second item entry into the first item entry.'" (emphasis added)

Applicant respectfully disagrees for several reasons. To begin with, the Examiner's Answer is understood to contend that "compression" occurs when "two contents" in different rows "[have] a single value or data which represents the articles in common." Applicant respectfully submits that this definition is not reasonable. Indeed, it is at odds with the usage of the term "compression" in both Applicant's specification and in Chaudhuri. For example, Chaudhuri describes that the "uncompressed hash table 133" shown in FIG. 4 with columns c_1 and c_2 "have several hash values in common." Thus, according to Chaudhuri, the mere fact that columns have values in common does not mean that they are "compressed."

Indeed, the definition of compression proposed in the Examiner's Answer would seem to encompass situations where the size of a data table were *increased*. For example, if every cell in a data table were duplicated, the duplicate cells would "have values or data in common" with the original cells. Thus, by the definition of compression set forth in the Examiner's Answer, these duplicate cells would be "compressed."

Accordingly, Applicant submits that the interpretation of the term "compression" relied upon by the Examiner's Answer is incorrect and unreasonable. Claims are to be interpreted as they would by those of ordinary skill in the art. Both the specification and the cited Chaudhuri reference indicate that those of ordinary skill do not consider multiple entries to be compressed merely because they "have values or data in common." Accordingly, an anticipation rejection cannot be supported on this basis. If the Examiner persists in the contention that "compression" occurs when "two contents" "[have] a single value or data which represents the articles in common," applicant respectfully requests that some iota of evidentiary support for this contention be provided.

At page 13, para. 3, the Examiner's Answer contends that

> "Applicant's allegation that only entries that have identical hash values are compressed into each other is clearly false in view of Chaudhuri's disclosure. In fact, entries that have different hash values (V2 and V3 are two different hash values) but with identical listing of articles associated with them (V2 is associated with C1 and C2, and V3 is also associated with C1 and C2) are compressed into each other."

Applicant respectfully disagrees. Once again, for the sake of convenience, Chaudhuri's FIGS. 4 and 5 are now reproduced.

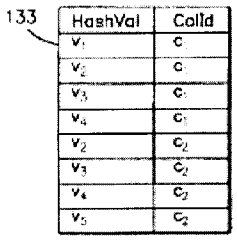

Fig.4

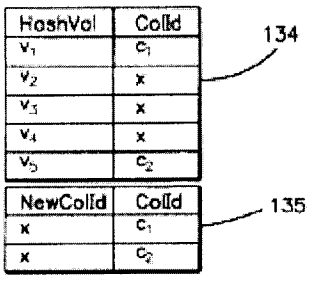

Fig.5

FIG. 4 is an example of a symbol table, and FIG. 5 is an example of a compressed version of the symbol table of FIG. 4. *See, e.g., Chaudhuri*, paras. [0014], [0015]. As can be seen, separate entries for V2 and V3 are maintained in FIG. 5. The entries for V2 and V3 are not compressed into each other. If the Examiner persists in the contention that the entries for V2 and V3 are somehow "compressed" in Chaudhuri's FIG. 5, applicant respectfully requests that some iota of support for the reasonableness of this contention be provided.

At page 14, line 4-6, in contending that Applicant is mistaken in the contention that none of the cited references teaches or suggests determining a cost-benefit ratio for the compression of item entries, the Examiner's Answer contends that:

> "Pugh clearly and rather explicitly discloses the concept of determining [a] cost-benefit ratio on column 2 line 50-52, column 3 lines 23-30. Pugh states that his method comprises determining the benefits and costs of reorganizing one or more objects based on the benefits and costs, thereby reorganizing objects having a higher benefit and correspondingly lower cost before reorganizing objects having a lower benefit and a correspondingly higher cost."

While this contention is true as far as it goes, Pugh's reorganization of a database object is not a compression of entries in an inverted index. Instead, Pugh's reorganization of a database "will generally eliminate wasted blocks in the [database] object, reduce the number of chained rows, and reduce the number of extents allocated. The free space distribution in a tablespace is improved by, for example, causing free space to be organized into fewer, larger extents as opposed to many, smaller extents." *See, e.g., Pugh*, col. 10, line 43-51. *See also Pugh*, col. 1, line 47-65. Pugh makes no mention whatsoever of database reorganization involving the compression of entries in an inverted index, but instead discusses the elimination of

fragmentation and chaining. Thus, as discussed in the Appeal Brief, Pugh has nothing to do with the compression of entries in an inverted index and also does not recognize any cost associated with such a compression.

If the Examiner persists in the contention that Pugh somehow involves compression of entries in an inverted index, applicant respectfully requests that some iota of evidentiary support for this contention be provided.

At page 14, line 11-19, in contending that the combination of Chaudhuri, Pugh, and Spencer is proper, the Examiner's answer contends that one of ordinary skill would find it obvious to add features from Pugh and Spencer to Chaudhuri since the features "add significant benefits."

Applicant respectfully disagrees. As discussed in the Appeal Brief and above, Pugh relates to the reorganization of a database. There is no reason to believe that features from Pugh are relevant to Chaudhuri at all, much less would "add significant benefits."

Spencer's primary concern is with efficient information retrieval from a database. *See, e.g., Spencer*, col. 1, line 46-49. Spencer describes that a static cache of "important" documents can be used to supplement searches of an inverted index. *See, e.g., id.*, col. 1, line 49-53. *See also id.*, col. 3, line 33-52. The static cache can be a primary index of the inverted index. *See, e.g., id.*, col. 4, line 20-21. The static cache contains an entry for each term of the inverted index that exceeds a certain number of document tuples. *See, e.g., id.*, col. 3, line 36-38. The static

cache thus does not compress entries in the inverted index, but rather includes only a subset of the entries in the inverted index. There is no reason to believe that features from Spencer are relevant to Chaudhuri at all, much less would "add significant benefits."

At page 14, line 19-27, the Examiner's answer contends that Pugh, Spencer, and Chaudhuri are from "the same field of endeavor, namely saving storage or memory capacity in a database."

Applicant respectfully disagrees. As discussed above, Spencer primarily relates to information retrieval and describes that a static cache of "important" documents can be used to supplement searches of an inverted index. Such a static cache is *supplemental* and hence *increases* data storage requirements.

For these reasons, and the reasons stated in the Appeal Brief, Applicant submits that the final rejection should be reversed.

Please apply any charges or credits to Deposit Account No. 06-1050.

Respectfully submitted,

Date: March 7, 2008

John F. Conroy
Reg. No. 45,485

Fish & Richardson P.C.
12390 El Camino Real
San Diego, California 92130
Telephone: (858) 678-5070
Facsimile: (858) 678-5099
JFC/jhg
10808124.doc